EPA Responses to Peer Review Comments on the Validation of the 15-Day Intact Adult Male Rat Assay

Drafted: October 30, 2007

The following responses are provided by the EPA with the expectation to clarify several main topics that seemed to be the primary concerns raised by some or all of the peer reviewers who participated in the review of the 15-Day Intact Adult Male Rat Assay.

Main topics to be addressed:

1) Performance and quality control (QC) aspects of the hormonal assays
2) Data integrity
3) Data interpretation
4) Phytoestrogen content of feed and analysis
5) Positive and negative test chemicals

Performance and quality control (QC) aspects of the hormonal assays

The same hormonal assay kits from the same commercial suppliers were prescribed in the protocol for all laboratories as described in the ISR (Page 26, Lines 25-28). However, one laboratory used a different model estradiol kit (same supplier). Maintaining uniformity of the hormonal assay kits among laboratories during the inter-laboratory exercise was done to minimize variation among laboratories due, in part, to potential differences in reagents (assay buffers and primary and secondary antibodies) and format (solid versus liquid phase) associated with hormonal assay kits for a particular hormone from different suppliers. The EPA is not expected to recommend any one commercial supplier over another in the final standardized protocol for the adult male assay but that a set of external QC standards will be prescribed (see details below) in addition to the internal QC standards provided in each assay kit to ensure that contract laboratories can competently run the hormonal assays according to the manufacturer's criteria and that they are relevant for use with rat serum.

As described in the adult male ISR (Page 29, Lines 11-13), an attempt was made to have each laboratory prepare external QC standards for each hormonal assay. Apparently, the adult male assay protocol did not provide enough detail or clarity in this regard since each laboratory seemed to, partly or wholly, neglect the internal QC standard provided with each kit and prepare external QC standards in their place, which was done inconsistently across laboratories. The results, therefore, were of limited value for interpreting the performance of the hormone assays as was acknowledged in the ISR (Page 29, Lines 29-35) but invaluable towards optimizing the bioassay protocol. Hence, to first determine that a hormonal assay kit performed within the specifications provided by the manufacturer, the standardized adult male assay protocol will be revised so that the internal QC standards will be done in accord with the manufactures' recommendations.

Despite the general consideration that steroids (testosterone, DHT and estradiol) and thyroid hormones (T3 and T4) are not species specific, the protocol specified the preparation of external QC standards (ISR Appendix A) to provide additional information on the performance of the hormonal assays since these kits were initially developed for use with human serum. To determine whether the steroid and thyroid hormone assay kits could measure these hormones in rat serum effectively within a laboratory, a pool of charcoal-stripped rat serum was spiked with various concentrations of T3 and T4 as detailed in the individual laboratory reports for RTI and Charles River. The percent recoveries for the observed results were often within 10% of the expected results for all steroid and thyroid hormonal assays as documented in the individual laboratory reports.

More specifically, since there was special concern directed toward the use of human T3 and T4 assays with rat serum, the percent recovery of multiple concentrations of T3 and T4 in the pool of rat serum reported in the results by RTI and Charles River ranged from 105-115% and 90-110%, respectively. Additional evidence that supports the use of the thyroid hormonal assay kits with rat serum was the T3 and T4 results following phenobarbital treatment. Absolute and relative changes were highly consistent with the results in the vehicle-control group and treatment groups across laboratories for each dose level and in accord with toxicological and biological historical results (ISR Summary Tables 9 and 14, respectively). Furthermore, the manufacturer of the T3 and T4 assay kits provided technical in-house data where rat serum was spiked with multiple concentrations of T3 and T4 which resulted in similar ranges in percent recovery. Hence, the results in the ISR combined with the results reported in two general review articles, a review listing numerous research articles with the adult male assay and thyroid hormone results published in peer reviewed scientific journals, and veterinary application documents by DPC referenced below are just a few sources which support the general consensus that commercial T3 and T4 assay kits developed for use with human serum are relevant for use with rat serum.

Davies DT. 1993. Assessment of rodent thyroid endocrinology: Advantages and pit-falls. Comparative Haematology International 3:142-152.

Christian MS, Trenton NA. 2003. Evaluation of thyroid function in neonatal and adult rats: The neglected endocrine mode of action. Pure and Applied Chemistry 75:2055-2068.

O'Connor, J.C., Cook, J.C., Marty, M.S., Davis, L.G., Kaplan, A.M., and Carney, E.W. (2002c). Evaluation of Tier I screening approaches for detecting endocrine-active compounds (EACs). *Crit. Rev. Toxicol.* **32,** 521-549.

Siemens Medical Solutions Diagnostics (Diagnostic Products Corporation, DPC), Coat-A-Count TKT3 (total T3) and TKT4 (total T4), Veterinary application documents of T3 (March 26, 1993, ZV106 A) and T4 (March 24, 1993, ZV103 A).

Considering the inconsistency in preparing the external QC standards across laboratories as acknowledged in the ISR (Page 29, Lines 11-13), especially for the steroid and thyroid hormone assay kits, the adult male standardized protocol will be revised so that all laboratories will clearly

understand the basis and process for spiking a pool of rat serum with respective steroids and thyroid hormones to assess percent recovery and to do serial dilutions of rat serum to check for parallelism against the human standard curve and establish in-house reliability for using the human kits with rat serum. Once established and on record, it is not expected that a laboratory will have to repeat this in-house mini-validation (i.e., spike samples and run serial dilutions) of the assay kit unless an assay kit or model number is used that is different from the one that was validated.

Expectedly, there should have been a total number of 60 experimental serum samples that could be run in one assay for each of the hormones without a major drift in the results from the beginning to the end of the assay. As expected for most hormones in each of the laboratories, the samples were run in one assay except if there were samples that needed to be rerun because they were outside the limits of the reference standard curve. Nonetheless, the standardized protocol will be revised to emphasize that all experimental samples including reference standards and QC samples be run in duplicate (QC samples will also be run in replicates of duplicate samples at the beginning, middle and end of each assay) in one assay for each hormone to avoid between assay variation and to get a more accurate assessment of within assay variation.

There was some concern raised as to the purpose of some of the hormonal assays (e.g., estradiol and FSH) in the adult male rat assay. The entire suite of hormonal assays was run in the inter-laboratory validation, in part, as an opportunity to assess the consistency of the results of the various assays across laboratories and compare with historical results. As was noted in the ISR (Page 26, Lines 19-20), not all hormonal assays may be necessary for a particular test chemical. Considering that most organ endpoints in the adult male are either androgen dependent or related to the thyroid, a core group of hormone assays (e.g., testosterone, LH, TSH, and T4) will likely be recommended and that the other assays (e.g., FSH, prolactin, estradiol, T3 and DHT) could be done if needed to support or negate initial results. A revision of the protocol will be made to clarify this aspect accordingly.

Data integrity

An intra-laboratory statistical analysis plan was prescribed for each laboratory prior to the inter-laboratory validation exercise as detailed in Appendix B of the adult male ISR. This was done, in part, to provide uniformity in the way the results were tabulated and analyzed within each laboratory and to facilitate inter-laboratory statistical analyses. Standardization of the statistical approaches within laboratory was done to minimize the operational variation among laboratories due to different laboratories having different standard operating procedures for the tabulation and statistical analyses of the data.

There was a reviewer concern that some of the statistical analyses used in the inter-laboratory validation were done post hoc. As indicated above, the main statistical approaches within and among laboratories were done a priori. For comparative purposes, an additional analysis was done after the initial analysis of the inter-laboratory results as described in the ISR (Appendix B). The second analysis was not done to replace the first as is evident in respective tables in Appendix B where the probability results of both statistical approaches are presented. However, since the second analysis was judged to be more appropriate for a small number of comparisons

(n=3 laboratories), the inter-laboratory validation results are expressed with the probability values from the second analysis as indicated in the ISR (Page 28, Lines 13-14).

There was also a reviewer concern that there was a failure to inspect the data. That is, many of the values for the SEs were the same across all dose levels for a given endpoint. For clarification, there was a common statistical plan provided to each laboratory as was discussed above and is presented in each of the individual laboratory reports (RTI, Appendix 6; WIL Appendix, G; Charles River, Appendix 9). Briefly, the plan indicated that tests for heterogeneity of variance were to be carried out on the data. For each endpoint, the extent of heterogeneity of variability was assessed across treatment groups. A one-way analysis of variance model was fitted to the data, including the factor treatment (fixed). Three versions of the model were fitted to test for heterogeneity of residual variance:

1. Separate variances for each treatment group (7 variances)
2. Separate variances for each substance (or control) (3 variances)
3. Common variances across all groups

For each endpoint, these models were compared by likelihood ratio tests and a "best" model compatible with the data was adopted. Hence, there were three possible outcomes for each endpoint.

1. Different variances within each test compound (T; control, linuron, phenobarbital) and dose (D). Denoted as T*D.
2. Different variances within each test compound (T) but constant across doses within compound. Denoted as T.
3. Constant across test compounds and doses. Denoted as All.

WIL and Charles River laboratories presented in their reports (preceding the summary tables) the covariance structures that were selected for each endpoint. In general:

For the "T*D" covariance structure, the LS means were presented and the SEs were different across dose groups within endpoint.

For the "T" covariance structure, the LS means were presented and the SEs were different across dose groups within endpoint.

For the "All" covariance structure, the LS means were presented and the SEs were similar across dose groups if the number of observations were the same for each dose level within endpoint. If the number of observations was different for a dose level within an endpoint, then the SE was also different and not the same for all dose levels within that endpoint. Note, the covariance structure was "All" for the T3 results in the WIL laboratory.

In addition, the conduct and results of each study provided by RTI, WIL and Charles River laboratories were subject to inspection for quality assurance according to federal guidelines (GLP) and a quality assurance plan provided by the primary contractor as stated in the protocols for each laboratory. Moreover, none of the laboratories reported protocol or GLP deviations

indicating the data were in error. Thus, the summary results reported in the adult male ISR are considered correct according to the common statistical plan that was provided to each of the contract laboratories and in compliance with quality assurance measures.

Data interpretation

The basis for data interpretation with respect to screening chemicals for endocrine activity was initiated during the pre-validation phase in which seminal studies, the diet-restriction studies (O'Connor et al., 1999;2000 in ISR), determined the range in final body weight allowable for interpretation of endocrine-mediated effects and whether interpretation of organ weight changes should rely on absolute or relative (to final body weight) changes. Initially, hormonal changes were also expected to be primary endpoints within the bioassay. However, considering the extent of the variation associated with many of the serum hormone concentrations as a result of the inter-laboratory validation exercise, hormonal measurements were reconsidered to be secondary or supplemental endpoints within the bioassay to support organ weight and histological changes as described in the ISR (Section 3.4.2). In addition, the weight of confidence in the results of a particular hormone was determined, in part, by the hormone assay performance results (i.e., coefficients of variation associated with QC samples).

As noted in Section 3.4.2, hormonal changes alone are of insufficient weight within the bioassay to make a conclusion of whether test chemical exposure was or was not endocrine-mediated. However, it was not intended that the hormonal results should be totally ignored or discarded if they were the only significant results. The protocol will be revised to indicate that if the only results are significant hormonal changes in which confidence in the hormone assay kit is relatively high then the results should be considered in the context of the overall results within the Tier-1 screening battery as an indicator of a potential endocrine-mediated effect.

It was also not intended that significant effects on relative weight changes of the testes and epididymides and absolute weight changes of the prostate, seminal vesicles with coagulating gland and thyroid be ignored or discarded from consideration within the bioassay. The protocol will be revised to clarify that based on the diet-restriction studies, absolute weight changes of the testes and epididymides and relative weight changes of the prostate, seminal vesicles with coagulating gland and thyroid will be weighted more heavily than the relative and absolute weight changes of respective organs in determining the results within the adult male assay as well as using the results in the context of other assays in the Tier-1 screening battery.

Phytoestrogen content of feed and analysis

There was concern by some reviewers regarding the diet and its phytoestrogen content during the testing period. The same commercial source was used to supply the same feed low in phytoestrogen content (Taklad 2018) and, in some instances (RTI and WIL laboratories) the same lot number, to each of the laboratories to minimize source variation as was indicated in the ISR (Section 4.3) and, in more detail, in the individual laboratory reports. In the study protocol, it was specified that the phytoestrogen content of the diet not exceed 300 ug/g (ppm) based on levels recommended for the proposed Uterotropic screening assay. Analysis of phytoestrogen (i.e., genistein, daidzein and glycitein) content in feed samples before the studies started indicated the values were well below the phytoestrogen cap that was specified as was recorded in

the individual laboratory reports. Hence, the level of phytoestrogen content in the diet during testing was controlled similarly for each contract laboratory.

Apparently, there is no documented information on the carryover effect within animals on reproduction or thyroid function following a change in diet from relatively high phytoestrogens to a diet relatively low in phytoestrogens in intact adult male SD rats and seemingly no consensus whether phytoestrogens in the diet for about 2 weeks can alter terminal weights of primary and secondary sex organs and thyroid gland, histomorphology of the testes, epididymides and thyroid, and serum concentrations of reproductive steroids, gonadotropins and thyroid hormones in intact adult male SD rats. Moreover, there were no significant changes in target organ weights, histomorphology and hormone concentrations following treatment of intact adult male rats with a high dose (1000 mg/kg/d) of genistein for 15 days even though there was a significant decrease in final body weight indicative of exposure as reported in the ISR (Section 3.1.1). Nonetheless, since a rodent diet with relatively low levels of phytoestrogens is readily and commercially available, the protocol will be revised to recommend that a diet low in phytoestrogens (<300 ug/g) be considered or at least that phytoestrogen content be available or analyzed before the study starts.

Positive and negative test chemicals

The number of positive test chemicals in the inter-laboratory validation study was limited so that more contract laboratories could be added to statistically compare the results across laboratories. The two chemicals chosen for the inter-laboratory study were meant to challenge the main modes of endocrine action (androgen and thyroid) that the intact adult male was designed to cover. As is evident in the list of pre-validation test chemicals (ISR, Table 4), the intact adult male reportedly can cover other endocrine receptor and non-receptor modes of action, especially steroidogenesis. Another limitation was that the negative test chemical was run only in an industrial laboratory.

The EPA acknowledges the efforts made by industry to initially develop the adult male assay and make it possible for the Agency to support inter-laboratory validation and assay peer review. The EPA also recognizes continuing efforts being made by industry to further validate the assay with more positive and negative test chemicals in multiple independent contract laboratories to strengthen the assay.

The American Chemistry Council (ACC) is sponsoring a study in early 2008 with results available before August 2008. See the RfP included in SAP review package for details. In brief, emphasis is on more chemicals than laboratories. Two contract laboratories will use the standardized 15-day intact adult male rat assay protocol to run allyl alcohol (toxic negative), fradrazole (positive for steroidogenesis, specifically aromatase inhibition), and DE-71 and iopanoate (positive as thyroid toxicants). Although there are only two contract laboratories, a qualitative or semi-quantitative evaluation of the observed results will be compared between laboratories and with expected historical data to primarily determine reproducibility.